

## Research Paper

# Serum microRNA profiles as prognostic biomarkers for HBV-positive hepatocellular carcinoma

Hao-Tu Zhu<sup>1,\*</sup>, Abdulbaqi M. E. Hasan<sup>1,\*</sup>, Rong-Bin Liu<sup>1</sup>, Zi-Chen Zhang<sup>1</sup>, Xiao Zhang<sup>1</sup>, Jing Wang<sup>1</sup>, Hai-Yun Wang<sup>1</sup>, Fang Wang<sup>1</sup>, Jian-Yong Shao<sup>1</sup>

<sup>1</sup>State Key Laboratory of Oncology in South China, Collaborative Innovation Center for Cancer Medicine, and Department of Molecular Diagnostics, Sun Yat-sen University Cancer Center, Guangzhou, 510060, China

\*These authors contributed equally to this work

**Correspondence to:** Jian-Yong Shao, email: shaojy@sysucc.org.cn

**Keywords:** serum microRNA, biomarkers, hepatocellular carcinoma, prognosis, deep sequencing

**Received:** January 28, 2016

**Accepted:** May 28, 2016

**Published:** June 15, 2016

## ABSTRACT

To establish serum microRNA profiles as prognostic biomarkers in hepatocellular carcinoma patients (HCCs), we used deep sequencing to screen serum microRNAs in a discovery set. Twelve up-regulated serum miRNAs were selected for qPCR analysis in a training set. miR-192-5p and miR-29a-3p were identified and associated with HCC prognosis. HCCs with high concentrations of miR-192-5p and miR-29a-3p had poorer overall survival (OS) and progression-free survival (PFS) than those with low concentrations. We calculated a prognostic index (PI) score and classified patients into low-, medium- and high-risk groups. OS and PFS among the 3 groups from the training set were significantly different (all  $P < 0.05$ ). PI (PI<sub>OS</sub>, PI<sub>PFS</sub>) score was the only independent prognostic predictor for OS and PFS of HCCs in the training set. These results were further confirmed in a validation set. In conclusion, differentially expressed serum miRNAs can be helpful for predicting survival in HCCs.

## INTRODUCTION

Hepatocellular carcinoma (HCC) is the most common and highly malignant hepatoma, the fifth most common cancer, and the third most common cause of cancer-related death in the world [1, 2]. Such a high fatality rate shows the unsatisfactory outcome of HCC and the lack of effective therapeutic strategies for this disease [3], mainly due to frequent postsurgical recurrence and metastasis [4, 5]. Discovery of specific biomarkers for accurate classification and stratification of HCC would be important in improving the prognostic assessment of HCC patients.

The discovery of microRNAs (miRNAs) has opened new avenues for cancer diagnosis, prognosis and prediction of treatment response [6]. miRNA signatures in HCC tissue have been shown to be associated with patient survival [7–9]. miRNAs also circulate in the blood in a cell-free form, stabilised by incorporation into microvesicles or RNA-binding proteins such as Ago2 [10–14]. miRNAs obtained from serum have been reported to provide prognostic information in patients with non-small-cell lung cancer and pancreatic cancer, indicating that serum miRNAs can indeed serve as prognostic markers in cancer patients [15–17].

In fact, differential expression of several miRNAs in the serum, including miR-16, miR-122, miR-21, miR-223, miR-25, miR-375 and let-7f in patients with HCC, patients with hepatitis B, and healthy individuals has been reported recently [18, 19]. However, most of those studies are confined to diagnostic miRNA signatures, and little research has been done to evaluate the prognostic value of circulating miRNAs in HCC. Thus, investigation of aberrant circulating miRNAs as prognostic biomarkers for HCC is an attractive and promising field.

Here, we describe the differential expression of serum miRNAs in 50 hepatitis B virus (HBV)-positive HCCs and 50 healthy controls (HCs) using deep sequencing technology in a discovery set. We used these results to establish a risk model in a training set of 74 HBV-positive HCCs and further validated the risk model in a validation set of 100 HBV-positive HCCs.

## RESULTS

### High-throughput sequencing of small RNAs from HCCs and HCs

To derive optimal serum miRNA profiles, we carried out high-throughput next-generation sequencing of 50

HCC samples and 50 HC samples in the discovery stage. All annotated small RNAs are explained in Supplementary Figure S1. A total of 12 993 179 and 13 117 991 reads were sequenced from 2 small RNA libraries constructed from HC and HCC samples, respectively. From HCs, 11 324 354 clean reads were obtained, which accounted for 87.2% of reads, whereas 11 667 958 clean reads were obtained for HCC samples. We also sought to obtain the sequences of small RNAs shared between the 2 groups (see Supplementary Figure S2A and Supplementary S2B).

### Differential expression of serum miRNAs screened by sequencing in the discovery set

A total of 146 known human miRNAs were found by use of the Illumina/HiSeq2000 platform at the discovery stage (see Supplementary Table S1). Following BLASTN searches and further sequence analysis, 13 miRNAs were found that were unique to HCCs and 64 miRNAs for HCs. In the serum of HCCs and HCs, 69 miRNAs coexisted (Figure 1A and 1B) that had previously been deposited in the miRBase database. On the basis of our criteria (miRNA read counts > 100, sequencing-read count ratio > 8 and *P* values < 0.05), 12 up-regulated and 7 down-regulated miRNAs with significantly different expression levels between HCCs and HCs were identified (Figure 1C). Because we were interested in potential markers that were practical and convenient for predicting HCC prognosis in

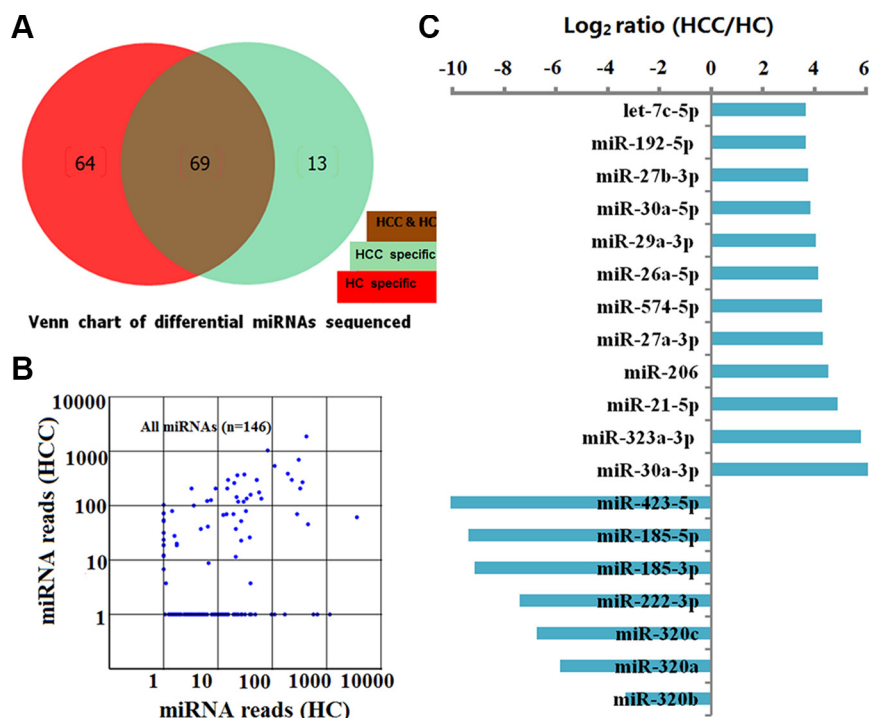
clinical applications, ultimately we selected 12 miRNAs up-regulated in HCC for further consideration in the training set.

### Correlation of differentially altered serum miRNAs and HCCs prognosis in the training set

To validate the sequencing data, we performed qPCR for each of these 12 up-regulated miRNAs screened by sequencing, and found that only 8 candidate miRNAs (let-7c-5p, miR-192-5p, miR-21-5p, miR-27a-3p, miR-27b-3p, miR-29a-3p, miR-30a-5p and miR-574-5p) were differentially altered between 74 HCCs and 60 HCs in the training set (see in the Supplementary Figure S3). Kaplan–Meier analysis confirmed that miR-192-5p and miR-29a-3p were associated with HCCs prognosis and progression, while the remaining miRNAs were not found to be significantly predictive for HCCs outcome (Figure 2).

Specifically, HCCs in the training set with high concentrations of miR-192-5p had poorer OS (log-rank test, *P* = 0.001) and PFS (*P* = 0.001) than those with low concentrations of miR-192-5p (Figure 2A and 2B). Similarly, HCCs in the training set with high concentrations of miR-29a-3p had poorer OS (*P* = 0.001) and PFS (*P* = 0.023) than those with low concentrations of miR-29a-3p (Figure 2C and 2D).

A univariate Cox regression model indicated that 6 variables including AFP, Barcelona Clinic Liver Cancer



**Figure 1: Characterization of 146 known human miRNAs screened by sequencing between HCs and HCCs in the discovery set.** (A) Number and overlap of known human miRNAs between HCs (*n* = 50) and HCCs (*n* = 50). (B) Scatter plot of miRNA reads for 146 miRNAs sequenced from 50 HCs and 50 HCCs. Each dot represents 1 miRNA. (C) 12 up-regulated and 7 down-regulated miRNAs with significantly different expression levels between HCs and HCCs (miRNA read counts > 100, sequencing-read count ratio > 8 and *P* values < 0.05).

stage (BCLC stage), tumor size, vascular invasion, miR-192-5p and miR-29a-3p were correlated with the survival of HCCs (see Supplementary Table S2). Subsequently, multivariate analyses revealed that BCLC stage and miR-29a-3p were independently correlated with the OS of HCCs, while AFP, BCLC stage and miR-192-5p were independently correlated with the PFS of HCCs. Taken together, these findings suggest that AFP, BCLC stage, miR-29a-3p and miR-192-5p are independent prognostic factors for survival in the training set (Table 1).

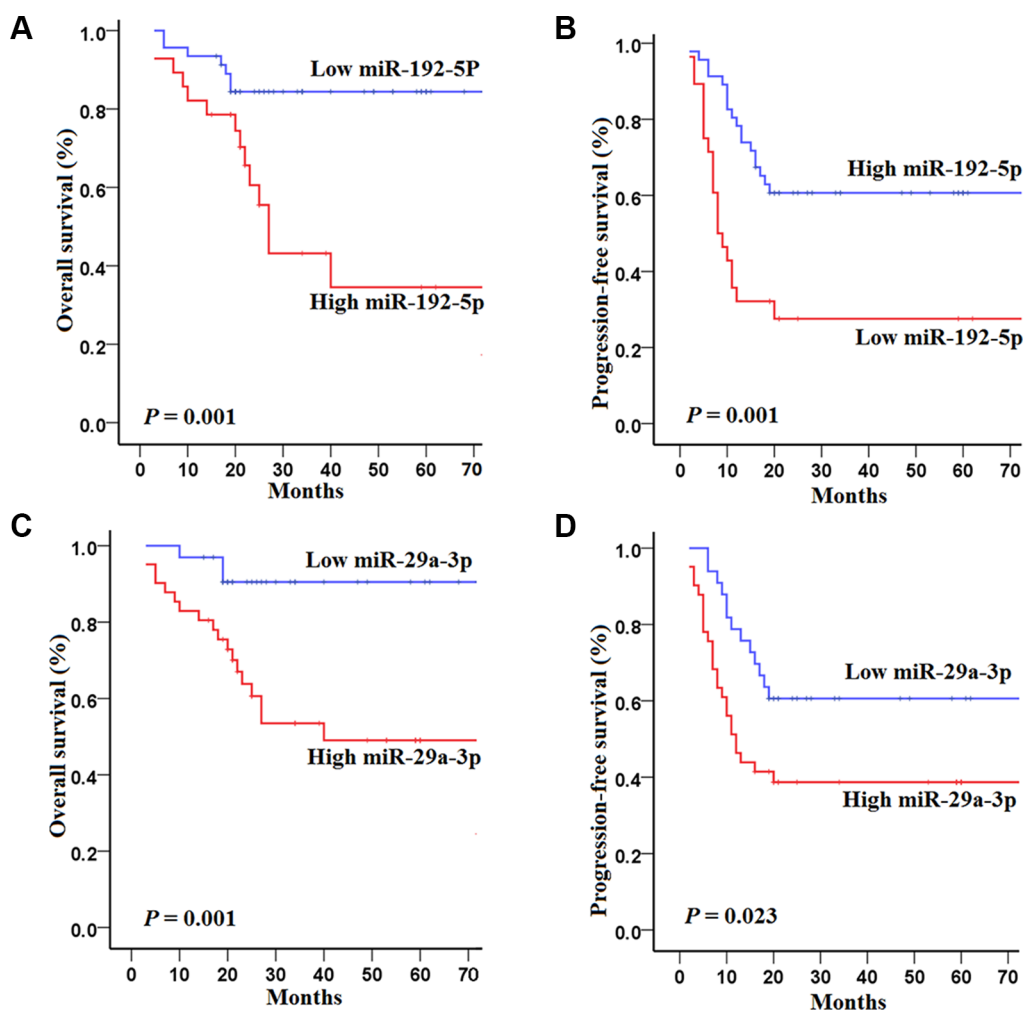
### Dynamic changes of serum miRNAs among HCCs, cirrhosis, and HCs in the training set

Four of the 8 miRNAs had significantly different expression levels between the HCCs and control groups (healthy and cirrhosis) in the training set, as shown in Table 2. These were let-7c, miR-27b-3p, miR-29a-3p, and miR-192-5p. With respect to the diagnostic value of the

2 miRNAs, a ROC curve was used to distinguish HCCs from HCs in the training set. The AUCs for HCC diagnosis determined from miR-192-5p and miR-29a-3p were 0.69 (95%CI 0.60–0.76,  $P = 0.0001$ ) and 0.71 (95%CI 0.62–0.78,  $P < 0.0001$ ), respectively (Figure 3A–3B).

The median levels of the 2 serum miRNAs were significantly higher in HCCs than in control groups (healthy controls and cirrhosis patients) (Figure 3C–3D). The median levels of the 2 serum miRNAs were significantly higher in both early and late stage HCCs than in controls (Figure 3C–3D). In addition, miR-192-5p showed differences in median concentration in between early and late stage HCCs (Figure 3C). These findings indicate that miR-192-5p and miR-29a-3p have good value for early event detection, specifically for miR-192-5p in predicting tumor recurrence and metastasis.

In HCC patients, a decrease of the 2 serum miRNAs concentration between pre- and post-operated HCCs was observed in HCCs serum in the training set (Figure 4A–4B).



**Figure 2: Kaplan–Meier curves of overall survival (OS) and progression-free survival (PFS) for 74 HCC patients in the training set with high and low concentrations of serum miRNAs. (A–B) OS and PFS for 74 HCCs in the training set according to the concentrations of miR-192-5p; (C–D) OS and PFS for 74 HCCs in the training set according to the concentrations of miR-29a-3p. The survival rates were compared using log-rank test.**

**Table 1: Multivariate Cox regression analyses of OS and PFS according to clinical variables and the levels of 2 serum miRNAs in the training set**

OS			
Variable	HR <sup>1</sup>	95% CI	P
BCLC stage <sup>2</sup>			
(0-A) <sup>3</sup>	1	reference	
(B-C) <sup>4</sup>	5.9	(1.8-18.8)	0.003
miR-29a-3p			
≤ 1.37	1	reference	
> 1.37	4.0	(1.2-13.9)	0.027
PFS			
Variable	HR	95% CI	P
AFP <sup>5</sup>			
≤ 35 ng/ml	1	reference	
> 35 ng/ml	2.1	(1.1-4.2)	0.031
BCLC stage			
(0-A)	1	reference	
(B-C)	5.7	(2.8-11.7)	0.000
miR-192-5p			
≤ 2.24	1	reference	
> 2.24	2.2	(1.1-4.2)	0.023

HR, hazard ratio; BCLC stage, Barcelona Clinic Liver Cancer stage; (0-A), Early stage; (B-C), Late stage; AFP, alpha fetoprotein.

**Table 2: Four of the 8 candidate miRNAs had significantly different expression levels between the HCC and control groups<sup>1</sup> in the training set**

NO.	miR name	HCC vs Control <sup>1</sup>	HCC vs healthy	fold change	HCC vs cirrhosis	fold change
		P value	P value		P value	
1	<b>let-7c-5p</b>	<b>0.0461</b>	0.035	1.491	0.119	1.312
2	miR-21-5p	0.0683	0.109	1.212	0.162	1.353
3	<b>miR-27b-3p</b>	<b>0.0001</b>	0.0002	1.910	0.0001	2.328
4	miR-27a-3p	0.2072	0.045	1.386	0.103	1.117
5	<b>miR-29a-3p</b>	<b>0.0003</b>	0.031	1.81	0.021	1.901
6	miR-30a-5p	0.2181	0.155	1.112	0.179	1.079
7	<b>miR-192-5p</b>	<b>0.0001</b>	0.0003	1.711	0.0002	1.973
8	miR-574-5p	0.2114	0.391	1.012	0.172	1.231

Control group includes healthy participants, and patients with cirrhosis.

### Risk model for predicting survival in HCCs from the training set

Finally, 3 variables, including BCLC stage, miR-192-5p and miR-29a-3p, were analyzed in a multivariate risk model to calculate the PI<sub>OS</sub> score, while another model combining BCLC stage, miR-192-5p and AFP was analyzed to calculate the PI<sub>PFS</sub> score. The scores were calculated by the following equations (method described

in Statistical analysis and Supplementary Materials): PI<sub>OS</sub> = 2.19 (0 or 1) + 1.45 (0 or 1) + 1.81(0 or 1); PI<sub>PFS</sub> = 1.89 (0 or 1) + 1.07 (0 or 1) + 0.69 (0 or 1). The PI<sub>OS</sub> and the PI<sub>PFS</sub> score had a range of 0-5.45 and 0-3.65, respectively. Consequently, patients with low-, medium- and high-risk had OS/PFS-predictor scores of < 1.81/0.69, 1.81-3.64/0.69-1.89 and > 3.64/1.89.

HCCs in the low-risk group had significantly longer OS (log-rank test, P = 0.000; P = 0.000) and PFS

( $P = 0.004$ ;  $P = 0.000$ ) than those in the medium- and high-risk group, and HCCs in the medium-risk group had longer OS ( $P = 0.038$ ) and PFS ( $P = 0.009$ ) than those in the high-risk group (Figure 5A–5B). In addition, ROC analyses revealed that the PI showed better survival prediction than the other prognostic characteristics as  $PI_{OS}$  and  $PI_{PFS}$  had the best AUCs. (all  $P < 0.05$ ) (Figure 5C–5D). Multivariate analyses also demonstrated that PI ( $PI_{OS}$  and  $PI_{PFS}$ ) was the only significantly independent prognostic predictor for OS (HR = 5.2, 95% CI 2.6–10.9;  $P = 0.000$ ) and PFS (HR = 3.1, 95% CI 2.0–4.7;  $P = 0.000$ ) of HCCs in the training set.

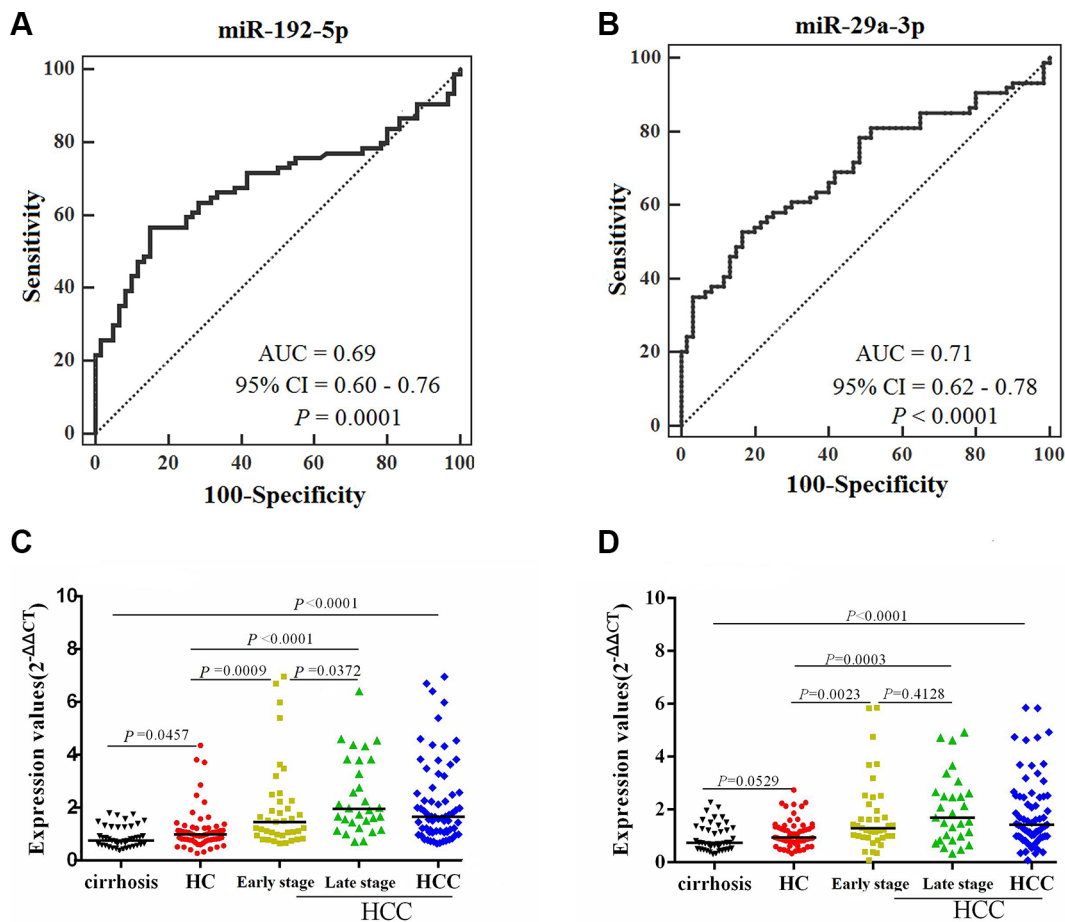
### Validating the risk model of PI ( $PI_{OS}$ , $PI_{PFS}$ ) in the validation set

To validate the prognostic value of the PI score, we quantified the levels of serum miR-192-5p and miR-29a-3p in a validation set of 100 HCCs and 70 HCs samples. We then calculated the risk score for each patient and classified patients into high-, medium- and low-risk groups using

the same formulas and method as used in the training set. Similar to what was seen in the training set, HCCs in the low-risk group had significantly longer OS (log-rank test,  $P = 0.000$ ;  $P = 0.000$ ) and PFS ( $P = 0.001$ ;  $P = 0.000$ ) than those in the medium- and high-risk group, and HCCs in the medium-risk group had longer OS ( $P = 0.016$ ) and PFS ( $P = 0.004$ ) than those in the high-risk group (Figure 6A–6B). Likewise, the  $PI_{OS}$  and  $PI_{PFS}$  had the best AUCs among prognostic factors in the validation set, showed a better survival prediction than the other prognostic characteristics by ROC analyses (all  $P < 0.05$ ) (Figure 6C–6D), and were independent prognostic predictors for OS (HR = 3.8, 95% CI 2.3–6.4;  $P = 0.000$ ) and PFS (HR = 2.7, 95% CI 2.0–3.7;  $P = 0.000$ ) of HCCs by multivariate analyses.

### The 2 miRNAs are overexpressed in HCC tissues

In HCC patients, an increase in miR-192-5p and miR-29a-3p staining intensity was observed in HCC tissues compared with corresponding normal adjacent tissues by FISH (Figure 7).

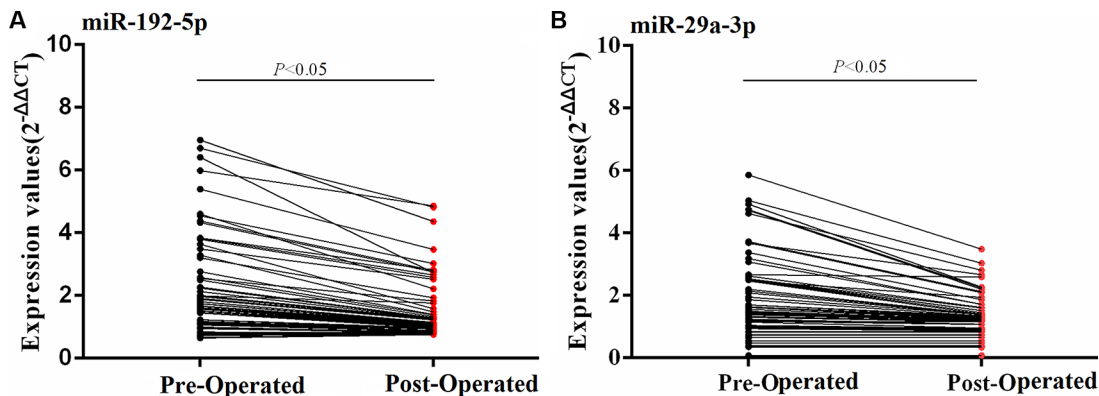


**Figure 3: Comparison of the sensitivity and specificity for HCC diagnosis by 2 miRNAs, with scatter plots representing dynamic expression changes of serum miRNAs between controls and HCCs.** AUCs of miR-192-5p (A) and miR-29a-3p (B) were analyzed with ROC curves indicating the potential value of the 2 miRNAs to identify HCC. Mann-Whitney unpaired tests using independent samples were used to compare serum miRNA concentrations of miR-192-5p (C) and miR-29a-3p (D) among the groups of cirrhosis patients ( $n = 43$ ), HC ( $n = 60$ ), early- ( $n = 42$ ) and late-stage HCC ( $n = 32$ ), respectively.

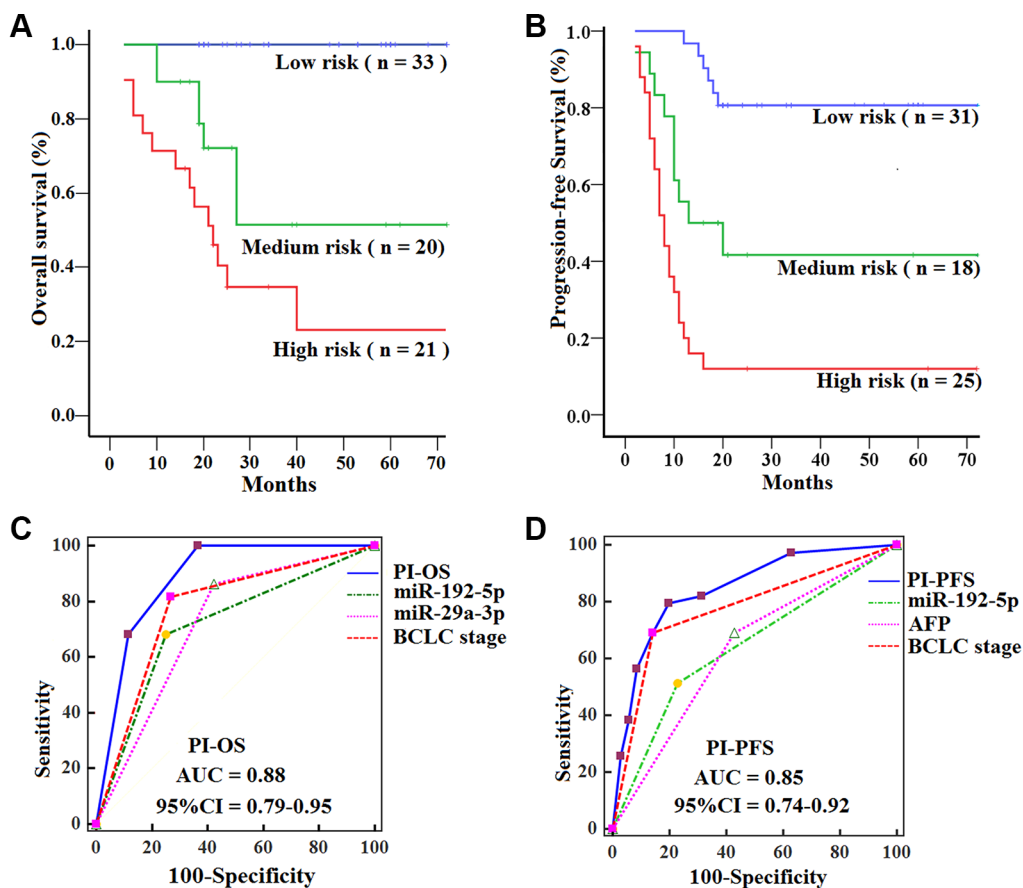
## DISCUSSION

We identified miR-192-5p and miR-29a-3p as inversely associated with survival and progression in HCCs. We found that a combination of miRNA expression levels and clinical variables, had a better prognostic value than BCLC stage alone.

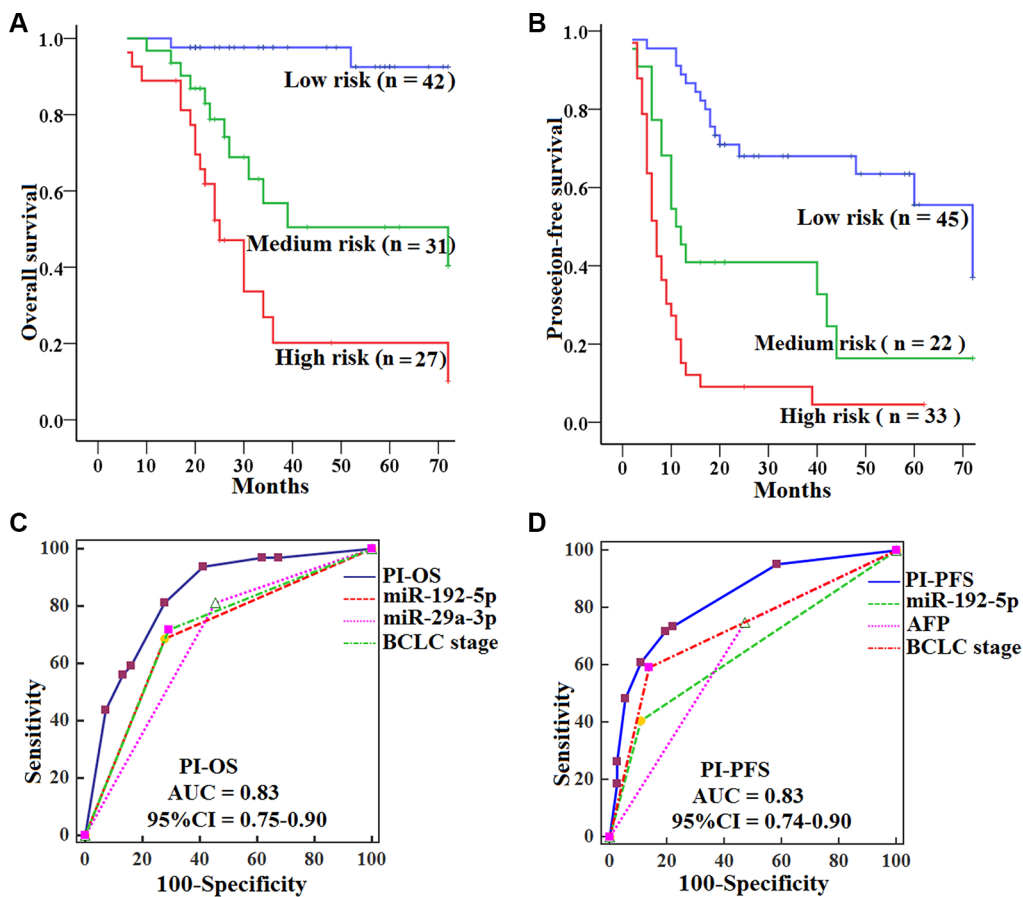
In the clinic, HCCs with the same stage who receive similar treatment regimens often show large variations in disease outcome, indicating that the current staging system may have reached its limitations for prognostic assessment. Staging in HCC patients is challenging because it requires staging and grading of both tumor and the chronic liver disease. Therefore, novel markers that reflect



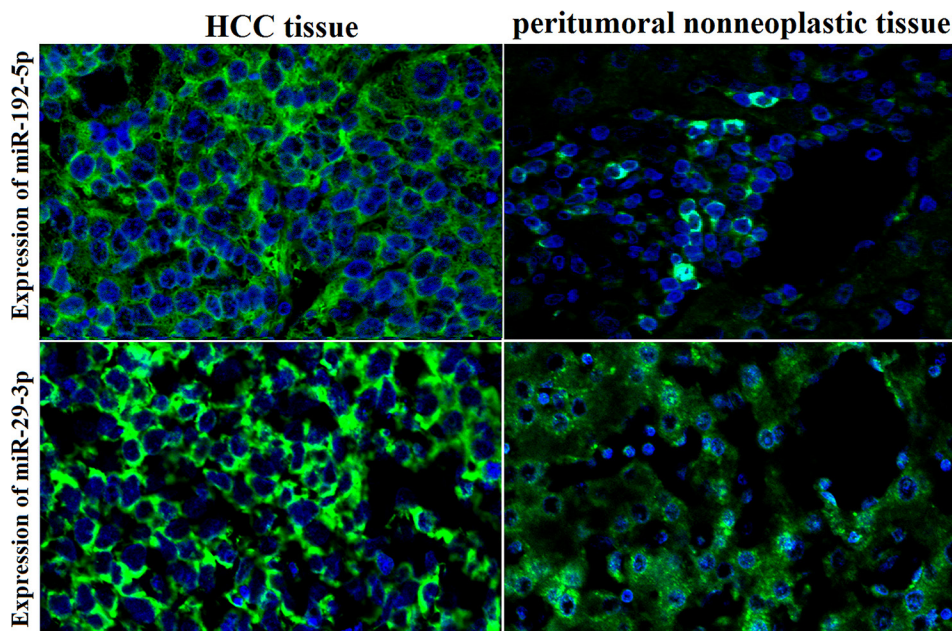
**Figure 4: Dynamic expression changes of serum miRNAs between pre- and post-operated HCCs in the training set.** The serum miRNAs concentration of pre- and post-operated HCCs were compared using Wilcoxon matched-pairs signed rank test.



**Figure 5: Kaplan-Meier curves of OS and PFS for HCCs with subtypes in the training set, with comparisons of the sensitivity and specificity for HCCs prediction of survival by  $PI_{OS}$ ,  $PI_{PFS}$ , clinical variables and miRNAs.** OS (A) and PFS (B) were calculated in 74 HCCs subdivided by PI into low-, medium- and high-risk subtypes. Log-rank test was used to calculate  $P$  value. AUCs of  $PI_{OS}$ ,  $PI_{PFS}$ , clinical variables and miRNAs for OS (C) and PFS (D) were analyzed with ROC curves.



**Figure 6:** Kaplan-Meier curves of OS and PFS for HCCs by subtypes in the validation set, with comparisons of the sensitivity and specificity for HCCs prediction of survival by  $PI_{OS}$ ,  $PI_{PFS}$ , clinical variables and miRNAs. OS (A) and PFS (B) were calculated in 100 HCCs subdivided by PI into low-, medium- and high-risk subtypes. Log-rank test was used to calculate *P* value. AUCs of  $PI_{OS}$ ,  $PI_{PFS}$ , clinical variables and miRNAs for OS (C) and PFS (D) were analyzed with ROC curves.



**Figure 7:** High expression of miR-192-5p and miR-29a-3p in HCC tissues by FISH. An increase in miR-192-5p and miR-29a-3p staining intensity was observed in HCC tissues compared with corresponding normal adjacent tissues by FISH. Positive *in situ* hybridization signals are visualized in green, while blue depicts DAPI nuclear stain ( $\times 400$  magnification).

the aggressiveness of the disease would complement the current set of prognostic and treatment algorithms.

Circulating miRNAs are stably detectable in serum or plasma [11, 23, 24] and have potential as novel prognostic biomarkers in a variety of cancers [15–17, 25–27]. We identified a risk model that was significantly associated with patient survival. Compared to patients with low- and medium-risk scores, HCCs with high-risk scores had reduced OS and PFS, verifying the prognostic value of serum miRNAs and the risk model of PI score.

Several publications suggest significant roles for miRNAs in the evaluation of cancer prognosis and diagnosis, such as lung cancer and HCC [28–30]. In the present study, serum miR-192-5p and miR-29a-3p as new prognostic parameters in HCCs are independent of classical BCLC stage. Therefore, we combined serum miRNAs and clinical variables to develop a risk model [15, 25–27, 30–34] of PI score, and the combination of serum miRNAs and BCLC stage had a better prognostic value than BCLC stage alone. So, the serum miRNA signature can add prognostic value to the BCLC stage system. As a result, HCCs could be stratified and distinguished into different risk groups, and could be treated with different approaches or intensities to improve their disease outcomes.

Logistic regression with penalized estimates may be used to develop prognostic models for binary outcomes, especially when limited data are available. Discrimination, calibration, and overall performance were taken into account in the construction of models [35]. ROC curve analysis was used to quantify a concordance statistic (c) in logistic regression [36]. The risk model was applied successfully to predict patient survival in lung cancer [30] and diffuse large B-cell lymphoma [31]. In our study, the PI divided HCCs into 3 groups correlated to differential survival. Such information not only may improve the predictive performance of the model, but also strengthen its clinical credibility.

Deep sequencing can be focused on both discovering novel miRNAs and acquiring a highly quantitative estimate of known individual miRNA species [37]. Developing minimally invasive methods including deep sequencing in the field of miRNAs for HCC diagnosis and prognosis is of great interest. Thus, a serum miRNA signature in HCCs, as shown in the present study, might be of great clinical interest as a routine testing procedure.

Decrease of both miR-192-5p and miR-29a-3p has been shown in HCC cells in several studies [8, 38–40], but two other studies have identified these same miRNAs as HCC diagnostic markers increased in the circulation of HCC patients [19, 41]. Our studies are in agreement with the literature on diagnostic circulating miRNA panels which indicate that miR-192-5p and miR-29a-3p have considerable clinical credibility for HCC diagnosis. In particular, miR-192-5p is expressed differentially in early and late stage HCC, and is a promising biomarker for HCC surveillance.

The American Association for the Study of Liver Diseases Practice Guidelines (July 2010) discarded AFP for surveillance and diagnosis. Therefore, there is a need for novel markers that would serve as a reliable method for early detection and postsurgical progression. This need might be well fulfilled with serum miR-192-5p and serum miR-29a-3p as alternative biomarkers which may be particularly helpful for AFP-negative HCCs

However, there are some limitations in our study. First, in the discovery stage, if we had compared the expression of serum miRNAs of the controls with that of the prognosis well and the poor HCC patients, it would be better in experimental design. Second, currently no established internal reference exists for circulating miRNAs. Non-human spike-in miRNAs, such as from *Caenorhabditis elegans* [42, 43] have been suggested as alternative references. Third, the two-miRNA signature was identified from Chinese patients at a single medical center. Fourth, the origins and mechanisms leading to the generation of serum miRNAs remain unclear. Finally, further studies are warranted to evaluate these miRNAs, to characterize the molecular mechanisms underlying miRNA function, and to discover potential therapy targets for HCC.

In conclusion, distinct serum miRNA profiles associated with survival exist in HCCs. Additionally, these serum miRNAs may be a robust predictor of prognosis for HCC, which may lead to more personalized therapy. Moreover, deep sequencing followed by qPCR validation may be a successful strategy to determine biomarkers in circulation.

## MATERIALS AND METHODS

### Study participants

All samples were obtained from confirmed cases of HCC at the time of diagnosis from the Sun Yat-sen University Cancer Center (SYSUCC) (Guangzhou, China). For the discovery stage of this study, we screened serum miRNAs in a set of 50 HCCs and 50 HCs consecutively obtained between January 2009 and April 2009. For subsequent study, (Training stage and Validation stage), we used randomly selected samples obtained between May 2009 and December 2013 from 174 HCCs, 43 cirrhosis patients and 130 HCs to test and validate miRNA expression at SYSUCC. HCs were individuals who came in for a routine physical examination and were found to be cancer-free. The characteristics of the 174 HCC patients are summarized in Table 3. The inclusion criteria for the study were a first time diagnosis of HCC with no history of other tumors and a follow-up time of > 1 year at the completion of this study. Informed consent was obtained for all participants. This study was approved by the medical research ethics committee of SYSUCC. All patients were treated with hepatic resection, local ablation or a combination of both.



**Table 3: Clinicopathological features and follow-up data for 174 patients in the training and validation data sets**

Variable	Training ( <i>n</i> = 74)	Validation ( <i>n</i> = 100)
	No.	No.
Age, years		
Median	52	51
Mean	51	51.5
Follow-up, months		
Median	27	28
Mean	35	36
Sex		
Male	63	84
Female	11	16
AFP <sup>1</sup>		
≤ 200 ng/ml	44	58
> 200 ng/ml	30	42
HBV-DNA		
≤ 1000 IU/ml	26	37
> 1000 IU/ml	48	63
Child-Pugh <sup>2</sup>		
A	66	88
B	8	12
Vascular invasion (+/-)		
+	8	12
-	66	88
Tumor size		
≤ 3 cm	28	35
> 3 cm	46	65
BCLC stage <sup>3</sup>		
0	15	17
A	27	40
B	24	31
C	8	12
Treatment regimen		
Resection	24	25
Local ablation <sup>4</sup>	25	30
Resection + Local ablation	25	45

AFP, alpha fetoprotein; Child-Pugh, a model to evaluate liver function;

BCLC, Barcelona Clinic Liver Cancer stage; 4. Including transarterial chemoembolisation (TACE) and radiofrequency ablation (RFA).

### Serum preparation and RNA extraction

Venous blood was incubated at room temperature for 1 hr, centrifuged at 500 g for 10 min and then 10,000 g for 30 min at 4°C to separate the serum. The supernatant

was transferred to fresh RNase-free tubes and stored at -80°C until use. We extracted total RNA using TRIzol<sup>®</sup>LS reagent (Life Technologies, Carlsbad, CA) following the manufacturer's recommendations. More detailed information is provided in the Supplementary Materials.

## Small RNA library construction and sequencing

We created 2 small RNA libraries according to the protocol published by Morin et al. [20]. To fully investigate the differences in serum miRNA profiles between HCCs and HCs in the discovery stage, sequencing was performed on the Illumina/HiSeq 2000 platform [Beijing Genomics Institute (BGI)]. More details are provided in the Supplementary Materials.

## Bioinformatics analysis of high-throughput data

Detailed information is provided in the Supplementary Materials.

## Analysis of differential miRNA expression

To identify miRNAs differentially expressed between HCCs and HCs in the discovery stage, we used the Bayesian method developed by Audic and Claverie [21]. miRNAs displaying at least a 2-fold change difference between the 2 groups were selected for further investigation. A detailed description of this procedure is shown in the Supplementary Materials.

## Mature miRNA qPCR

To find a miRNA signature with which to construct a risk model in the training set, SYBR<sup>®</sup>Green based quantitative PCR (qPCR) reactions were performed on the 7500 Real-time PCR systems (Applied Biosystems, Carlsbad, CA) to validate 12 mature up-regulated miRNAs screened by sequencing (see Supplementary Table S3). The miRNA signature was further confirmed in the validation set by qPCR method. The detailed method is described in the Supplementary Materials.

## Locked nucleic acid (LNA)-based *in situ* hybridization for miRNA in HCC

To study the spatial and temporal expression of miRNAs with high sensitivity and resolution, the miRNA fluorescein *in situ* hybridization (FISH) protocol were optimized and performed. More details are provided in the Supplementary Materials.

## Statistical analysis

To compare the sequencing data of the HCCs and HCs, we used the Wilcoxon–Mann–Whitney test. For the data obtained by qPCR, either the Mann–Whitney unpaired test or the Kruskal–Wallis test was used for the comparison between HCC and control (cirrhosis patients and healthy controls). We compared the serum miRNAs concentration of pre- and post-operated HCCs using Wilcoxon matched-pairs signed rank test. The levels of all miRNAs with significance values < 0.05 were considered statistically significant.

We defined overall survival (OS) and progression free survival (PFS) as the period between the date of diagnosis and recurrence/metastasis, death, or last day of follow-up. Survival curves were estimated by use of the Kaplan–Meier method with the log-rank test. We used a multivariate Cox regression analysis to investigate whether the selected miRNAs were independent prognostic factors of OS and PFS in HCCs.

For each miRNA expression profile, we plotted the sensitivity and specificity of each outcome as a predictor of death for HCC with receiver operating characteristic (ROC) curve analysis. We used ROC curves as an accuracy index for assessing the predictive power as well as evaluating the diagnostic performance of each variable. We used the ROC curves to select cutoff scores for dichotomizing each predictor according to the maximum area under the ROC curve (AUC). The cutoff scores were generated with MedCalc software, version 12.2.1.

Penalized maximum likelihood estimation was applied to select the optimal variables to establish a multivariate risk model for evaluating HCC patient outcome. To evaluate the joint efficacy of this multivariate risk model, we introduced a prognostic index (PI). In general, we defined the PI score in terms of the logistic regression model:  $PI = \text{logit}\{Y = 1 | X\} = \beta_0 + \sum \beta_i x_i$  [22], where  $Y$  is a

binary outcome variable (0 or 1),  $\beta_0$  is an intercept, and denotes the regression coefficients associated with the  $\beta_i$  design matrix  $X$  of covariables  $i$ . Specifically, in our study we calculated the PI score according to the formula:

$PI = \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3$ , where  $\beta$  is a variable assignment indicated in univariate Cox regression analyses for HCCs in the training set (see Supplementary Table S2). A ROC curve was generated for further analysis of the prognostic value of PI. We used the maximum Youden index to obtain optimal cutoff values for  $PI_{OS}$  and  $PI_{PFS}$  for prognostic assessment in HCCs. Subsequently, ROC curve analyses for each subgroup were applied to calculate the sensitivity and specificity of PI.

Two-sided  $P$  values of < 0.05 were considered to indicate statistical significance. Statistical analyses were carried out by use of SPSS 17.0.

## ACKNOWLEDGMENTS

The authors thank the local doctors and the patients who participated in our study.

## CONFLICTS OF INTEREST

The authors have no competing interests to declare.

## GRANT SUPPORT

This work was supported by grants from the National Natural Science Foundation of China (grant no. 81472522, 81272952), and a State Key Laboratory Grant at Sun Yat-sen University Cancer Center.

## REFERENCES

1. Parkin DM. Global cancer statistics in the year 2000. *Lancet Oncol.* 2001; 2:533–543.
2. Flores A, Marrero JA. Emerging trends in hepatocellular carcinoma: focus on diagnosis and therapeutics. *Clin Med Insights Oncol.* 2014; 8:71–76.
3. Schafer DF, Sorrell MF. Hepatocellular carcinoma. *Lancet.* 1999; 353:1253–1257.
4. Maluccio M, Covey A. Recent progress in understanding, diagnosing, and treating hepatocellular carcinoma. *CA Cancer J Clin.* 2012; 62:394–399.
5. European Association for the Study of the Liver, European Organisation for Research and Treatment of Cancer. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. *J Hepatol.* 2012; 56:908–943.
6. Ambros V. The functions of animal microRNAs. *Nature.* 2004; 431:350–355.
7. Coulouarn C, Factor VM, Andersen JB, Durkin ME, Thorgeirsson SS. Loss of miR-122 expression in liver cancer correlates with suppression of the hepatic phenotype and gain of metastatic properties. *Oncogene.* 2009; 28: 3526–3536.
8. Xiong Y, Fang JH, Yun JP, Yang J, Zhang Y, Jia WH, Zhuang SM. Effects of microRNA-29 on apoptosis, tumorigenicity, and prognosis of hepatocellular carcinoma. *Hepatology.* 2010; 51:836–845.
9. Ji J, Zhao L, Budhu A, Forgues M, Jia HL, Qin LX, Ye QH, Yu J, Shi X, Tang ZY, Wang XW. Let-7g targets collagen type I alpha2 and inhibits cell migration in hepatocellular carcinoma. *J Hepatol.* 2010; 52:690–697.
10. Lawrie CH, Gal S, Dunlop HM, Pushkaran B, Liggins AP, Pulford K, Banham AH, Pezzella F, Boultonwood J, Wainscoat JS, Hatton CS, Harris AL. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. *Br J Haematol.* 2008; 141:672–675.
11. Mitchell PS, Parkin RK, Kroh EM, Fritz BR, Wyman SK, Pogosova-Agadjanyan EL, Peterson A, Noteboom J, O'Brian KC, Allen A, Lin DW, Urban N, Drescher CW, et al. Circulating microRNAs as stable blood-based markers for cancer detection. *Proc Natl Acad Sci U S A.* 2008; 105:10513–10518.
12. Valadi H, Ekstrom K, Bossios A, Sjostrand M, Lee JJ, Lotvall JO. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat Cell Biol.* 2007; 9:654–659.
13. Arroyo JD, Chevillet JR, Kroh EM, Ruf IK, Pritchard CC, Gibson DF, Mitchell PS, Bennett CF, Pogosova-Agadjanyan EL, Stirewalt DL, Tait JF, Tewari M. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. *Proc Natl Acad Sci U S A.* 2011; 108:5003–5008.
14. Turchinovich A, Weiz L, Langheinze A, Burwinkel B. Characterization of extracellular circulating microRNA. *Nucleic Acids Res.* 2011; 39:7223–7233.
15. Hu Z, Chen X, Zhao Y, Tian T, Jin G, Shu Y, Chen Y, Xu L, Zen K, Zhang C, Shen H. Serum microRNA signatures identified in a genome-wide serum microRNA expression profiling predict survival of non-small-cell lung cancer. *J Clin Oncol.* 2010; 28:1721–1726.
16. Boeri M, Verri C, Conte D, Roz L, Modena P, Facchinetti F, Calabro E, Croce CM, Pastorino U, Sozzi G. MicroRNA signatures in tissues and plasma predict development and prognosis of computed tomography detected lung cancer. *Proc Natl Acad Sci U S A.* 2011; 108:3713–3718.
17. Liu R, Chen X, Du Y, Yao W, Shen L, Wang C, Hu Z, Zhuang R, Ning G, Zhang C, Yuan Y, Li Z, Zen K, et al. Serum microRNA expression profile as a biomarker in the diagnosis and prognosis of pancreatic cancer. *Clin Chem.* 2012; 58:610–618.
18. Xu J, Wu C, Che X, Wang L, Yu D, Zhang T, Huang L, Li H, Tan W, Wang C, Lin D. Circulating microRNAs, miR-21, miR-122, and miR-223, in patients with hepatocellular carcinoma or chronic hepatitis. *Mol Carcinog.* 2011; 50:136–142.
19. Zhou J, Yu L, Gao X, Hu J, Wang J, Dai Z, Wang JF, Zhang Z, Lu S, Huang X, Wang Z, Qiu S, Wang XY, et al. Plasma microRNA panel to diagnose hepatitis B virus-related hepatocellular carcinoma. *J Clin Oncol.* 2011; 29:4781–4788.
20. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu AL, Zhao Y, McDonald H, Zeng T, Hirst M, Eaves CJ, Marra MA. Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* 2008; 18:610–621.
21. Audic S, Claverie JM. The significance of digital gene expression profiles. *Genome Res.* 1997; 7:986–995.
22. Steyerberg EW, Eijkemans MJ, Harrell FE Jr, Habbema JD. Prognostic modelling with logistic regression analysis: a comparison of selection and estimation methods in small data sets. *Stat Med.* 2000; 19:1059–1079.
23. Chen X, Ba Y, Ma L, Cai X, Yin Y, Wang K, Guo J, Zhang Y, Chen J, Guo X, Li Q, Li X, Wang K, et al. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. *Cell Res.* 2008; 18:997–1006.
24. Gilad S, Meiri E, Yogev Y, Benjamin S, Lebanony D, Yerushalmi N, Benjamin H, Kushnir M, Cholakh H, Melamed N, Bentwich Z, Hod M, Goren Y, et al. Serum microRNAs are promising novel biomarkers. *PloS One.* 2008; 3:e3148.
25. Liu N, Cui RX, Sun Y, Guo R, Mao YP, Tang LL, Jiang W, Liu X, Cheng YK, He QM, Cho WC, Liu LZ, Ma J. A four-miRNA signature identified from genome-wide serum miRNA profiling predicts survival in patients with nasopharyngeal carcinoma. *Int J Cancer.* 2014; 134:1359–1368.

26. Kleivi Sahlberg K, Bottai G, Naume B, Burwinkel B, Calin GA, Borresen-Dale AL, Santarpia L. A serum microRNA signature predicts tumor relapse and survival in triple-negative breast cancer patients. *Clin Cancer Res*. 2015; 21:1207–1214.
27. Wang HY, Yan LX, Shao Q, Fu S, Zhang ZC, Ye W, Zeng YX, Shao JY. Profiling plasma microRNA in nasopharyngeal carcinoma with deep sequencing. *Clin Chem*. 2014; 60:773–782.
28. Karakatsanis A, Papaconstantinou I, Gazouli M, Lyberopoulou A, Polymeneas G, Voros D. Expression of microRNAs, miR-21, miR-31, miR-122, miR-145, miR-146a, miR-200c, miR-221, miR-222, and miR-223 in patients with hepatocellular carcinoma or intrahepatic cholangiocarcinoma and its prognostic significance. *Mol Carcinog*. 2013; 52:297–303.
29. Budhu A, Jia HL, Forgues M, Liu CG, Goldstein D, Lam A, Zanetti KA, Ye QH, Qin LX, Croce CM, Tang ZY, Wang XW. Identification of metastasis-related microRNAs in hepatocellular carcinoma. *Hepatology*. 2008; 47:897–907.
30. Yu SL, Chen HY, Chang GC, Chen CY, Chen HW, Singh S, Cheng CL, Yu CJ, Lee YC, Chen HS, Su TJ, Chiang CC, Li HN, et al. MicroRNA signature predicts survival and relapse in lung cancer. *Cancer Cell*. 2008; 13:48–57.
31. Lossos IS, Czerwinski DK, Alizadeh AA, Wechsler MA, Tibshirani R, Botstein D, Levy R. Prediction of survival in diffuse large-B-cell lymphoma based on the expression of six genes. *N Engl J Med*. 2004; 350:1828–1837.
32. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, Cheng CL, Wang CH, Terng HJ, Kao SF, Chan WK, Li HN, Liu CC, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007; 356:11–20.
33. Yang HI, Yuen MF, Chan HL, Han KH, Chen PJ, Kim DY, Ahn SH, Chen CJ, Wong VW, Seto WK, Group R-BW. Risk estimation for hepatocellular carcinoma in chronic hepatitis B (REACH-B): development and validation of a predictive score. *Lancet Oncol*. 2011; 12:568–574.
34. Liu N, Chen NY, Cui RX, Li WF, Li Y, Wei RR, Zhang MY, Sun Y, Huang BJ, Chen M, He QM, Jiang N, Chen L, et al. Prognostic value of a microRNA signature in nasopharyngeal carcinoma: a microRNA expression analysis. *Lancet Oncol*. 2012; 13:633–641.
35. Harrell FE Jr, Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. *Stat Med*. 1996; 15:361–387.
36. Harrell FE Jr, Califf RM, Pryor DB, Lee KL, Rosati RA. Evaluating the yield of medical tests. *JAMA*. 1982; 247:2543–2546.
37. Park G, Borkovich KA. Small RNA isolation and library construction for expression profiling of small RNAs from *Neurospora* and *Fusarium* using illumina high-throughput deep sequencing. *Methods Mol Biol*. 2012; 883:155–164.
38. Braconi C, Kogure T, Valeri N, Huang N, Nuovo G, Costinean S, Negrini M, Miotto E, Croce CM, Patel T. microRNA-29 can regulate expression of the long non-coding RNA gene MEG3 in hepatocellular cancer. *Oncogene*. 2011; 30:4750–4756.
39. Parpart S, Roessler S, Dong F, Rao V, Takai A, Ji J, Qin LX, Ye QH, Jia HL, Tang ZY, Wang XW. Modulation of miR-29 expression by alpha-fetoprotein is linked to the hepatocellular carcinoma epigenome. *Hepatology*. 2014; 60:872–883.
40. Yang YM, Lee WH, Lee CG, An J, Kim ES, Kim SH, Lee SK, Lee CH, Dhanasekaran DN, Moon A, Hwang S, Lee SJ, Park JW, et al. Galpha12 gep oncogene deregulation of p53-responsive microRNAs promotes epithelial-mesenchymal transition of hepatocellular carcinoma. *Oncogene*. 2015; 34:2910–2921.
41. Lin XJ, Chong Y, Guo ZW, Xie C, Yang XJ, Zhang Q, Li SP, Xiong Y, Yuan Y, Min J, Jia WH, Jie Y, Chen MS, et al. A serum microRNA classifier for early detection of hepatocellular carcinoma: a multicentre, retrospective, longitudinal biomarker identification study with a nested case-control study. *Lancet Oncol*. 2015; 16:804–815.
42. Toiyama Y, Takahashi M, Hur K, Nagasaka T, Tanaka K, Inoue Y, Kusunoki M, Boland CR, Goel A. Serum miR-21 as a diagnostic and prognostic biomarker in colorectal cancer. *J Natl Cancer Inst*. 2013; 105:849–859.
43. Shrivastava S, Petrone J, Steele R, Lauer GM, Di Bisceglie AM, Ray RB. Up-regulation of circulating miR-20a is correlated with hepatitis C virus-mediated liver disease progression. *Hepatology*. 2013; 58:863–871.